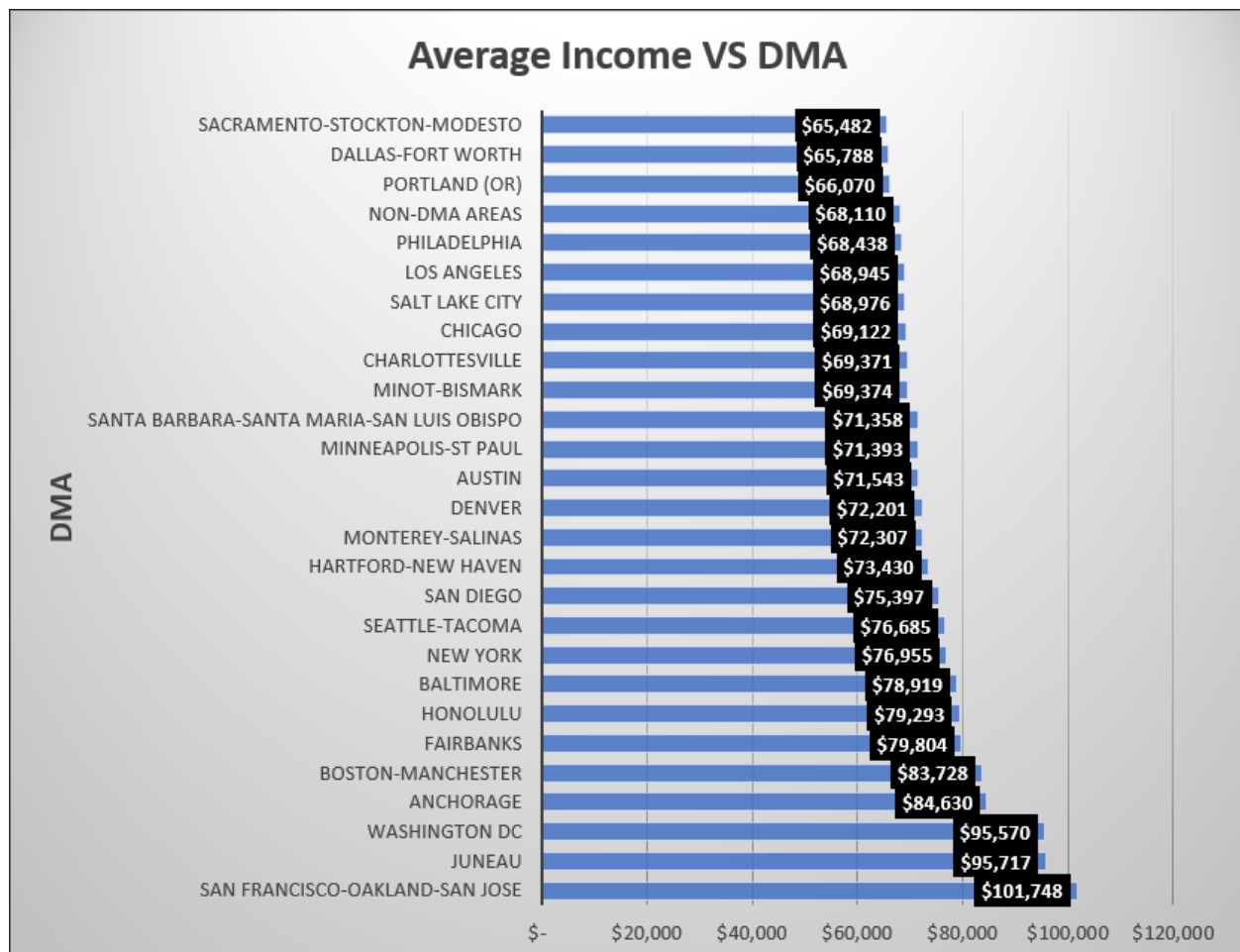


Analysis and Launch plan for Retail Startup

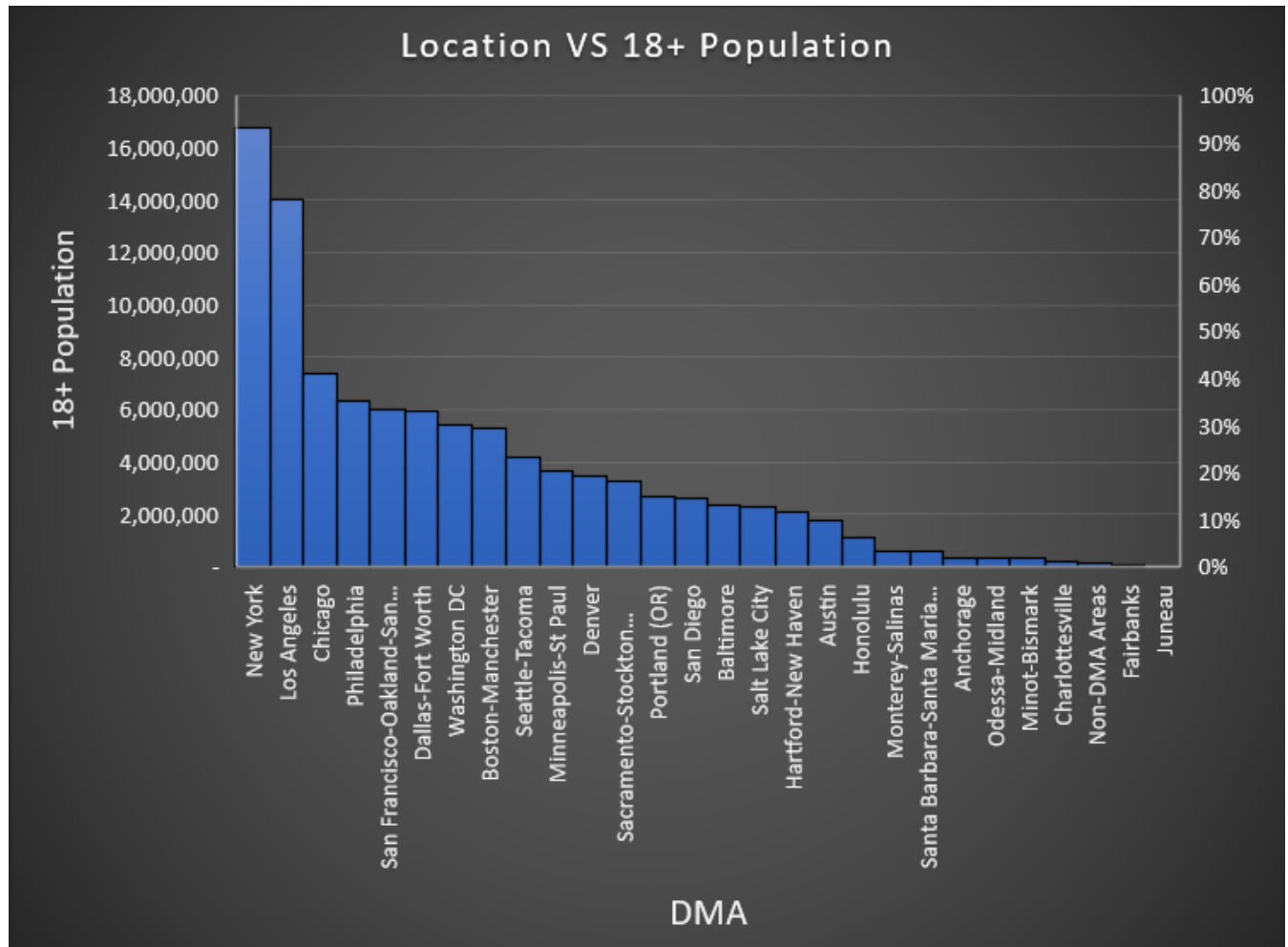
The goal of this analysis was to understand the current competitive landscape for nationwide multi-purpose home and grocery stores and propose 10 locations to launch a new “light” store. Prior to listing the 10 proposed locations, let us establish the baseline criteria for which these locations were selected.

The first criterion was the average household income of a given location. Based on the information provided by the U.S. Census Bureau and quoted on Investopedia, “What Is Middle Class Income? The Latest Numbers Available”, the average income of middle-income households is \$65,000. This figure is key in the analysis as it would serve very little purpose to open a store targeted for the middle income and higher in a location with an average household income below \$65K. This allows us to narrow down our findings to the following locations:

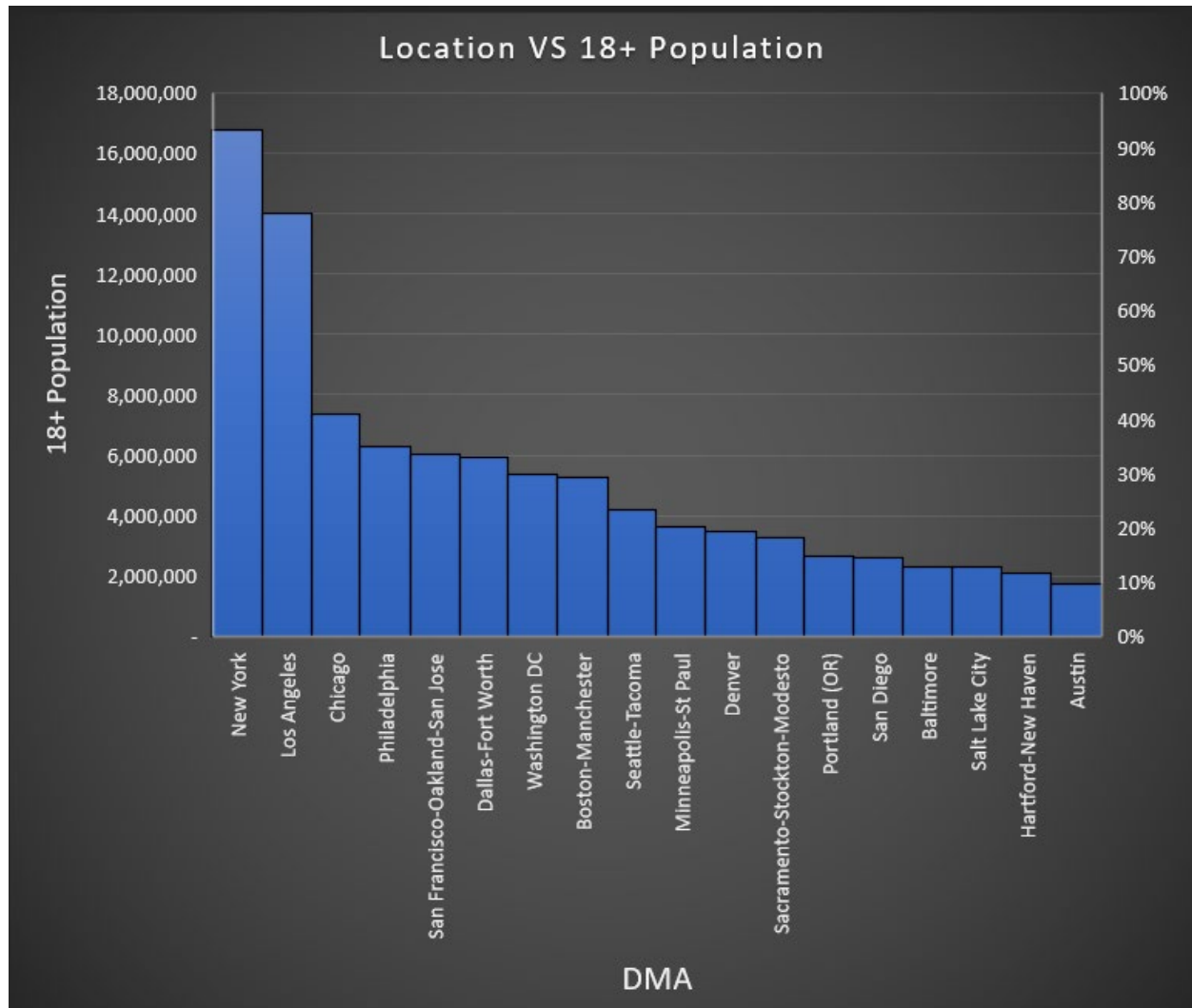


Based on the above graph we see that the list of possible locations to establish our startup is several- 27 excluding the non-DMA areas. However, something the above graph doesn't show is

the population for the locations listed. The below graph however does show the population of each location and provides some additional insight on what locations to immediately rule out due to smaller population count.



Locations from Honolulu and onward will be ignored as their population is either at ~1 million or below. This will in turn have our above graph look something like this:



Now that we have narrowed down our possible store locations we need to dive into the data from competitors. This data will help us understand how the 45 different competitor locations fair in their given locations versus the locations we were able to find above. The first step performed in the code was to merge our datasets and take a quick look at the different columns we have. The output from said action is found below:

```
Int64Index: 421570 entries, 0 to 421569
Data columns (total 16 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Store            421570 non-null  int64
1   Dept             421570 non-null  int64
2   Date             421570 non-null  datetime64[ns]
3   Weekly_Sales     421570 non-null  float64
4   IsHoliday        421570 non-null  bool
5   Type             421570 non-null  object
6   Size             421570 non-null  int64
7   DMA              421570 non-null  object
8   Temperature      265245 non-null  float64
9   Fuel_Price       265245 non-null  float64
10  CPI              265245 non-null  float64
11  Unemployment     265245 non-null  float64
12  Year             421570 non-null  int64
13  Month            421570 non-null  int64
14  Day              421570 non-null  int64
15  WeekOfYear       421570 non-null  float64
```

From this, we see that we have anywhere from 420K+ non-null entries to 265K+ non-null entries. This is a very healthy sized dataset to work with. Now that our data is set up for the additional work we will perform, let us take a look at the different DMA locations competitors are located in as well as how they perform based on their average weekly sales.



Based on the above distribution we can see some clear trends immediately. First being a large discrepancy between how much some competitor locations yield in average weekly sales versus others. Note that the values within each column are formatted as being in the thousands. I.e., 13.93 is 13930. The lowest locations average weekly sales were \$9.1K in Kansas City while the highest location- Houston- brought in \$22.83K. In addition, we can take an even closer look and

visualize how the different competitor locations have performed over the three years of total sales.



The trends we saw as a whole in our competitor data still hold even when magnified at an annual level alone. Do note however that there are some small differences between locations simply due to variances between years.

Now that we have a baseline of how our competitors performed versus the appropriate household income locations, a quick cross analysis between the two sets of locations will provide a rudimentary baseline for our 10 possible startup locations.

Competitor Locations	Income Based Locations	Match Status
Charlotte	Chicago	FALSE
Orlando-Daytona Brach-Melbourne FL	Denver	FALSE
	Salt Lake City	FALSE
	Austin	FALSE
Atlanta	New York	FALSE
Cleveland-Akron	San Francisco-Oakland-San Jose	FALSE
Dallas-Ft.Worth	Dallas-Fort Worth	FALSE
Houston	Boston-Manchester	FALSE
Kansas City	Seattle-Tacoma	FALSE
Tampa-St. Pete	Baltimore	FALSE
	Hartford-New Haven	FALSE
Austin	Los Angeles	TRUE
Chicago	Philadelphia	TRUE
San Diego	San Diego	TRUE
Denver	Washington DC	TRUE
Los Angeles	Minneapolis-St Paul	TRUE
Philadelphia	Sacramento-Stockton-Modesto	TRUE
Salt Lake City	Portland (OR)	TRUE

This quick analysis indicates that only 7 locations are shared between competitors' locations and appropriate income-based locations. Those being highlighted in green in the above figure. Those locations are all appropriate to select based on the income, population, and success of current competitors. Even when we reintroduce the eliminated locations due to population we have no additional locations to add to our 7 as seen below.

Competitor Locations	Income Based Locations	Match Status
Charlotte	New York	FALSE
Orlando-Daytona Brach-Melbourne FL	Los Angeles	FALSE
	Chicago	FALSE
	Philadelphia	FALSE
	San Francisco-Oakland-San Jose	FALSE
Atlanta	Dallas-Fort Worth	FALSE
Cleveland-Akron	Washington DC	FALSE
Dallas-Ft.Worth	Boston-Manchester	FALSE
Houston	Seattle-Tacoma	FALSE
Kansas City	Minneapolis-St Paul	FALSE
Tampa-St. Pete	Denver	FALSE
	Honolulu	FALSE
	Monterey-Salinas	FALSE
	Santa Barbara-Santa Maria-San Luis Obispo	FALSE
	Anchorage	FALSE
	Odessa-Midland	FALSE
	Minot-Bismark	FALSE
	Charlottesville	FALSE
	Non-DMA Areas	FALSE
	Fairbanks	FALSE
	Juneau	FALSE
	Sacramento-Stockton-Modesto	TRUE
Austin	Portland (OR)	TRUE
Chicago	San Diego	TRUE
San Diego	Baltimore	TRUE
Denver	Salt Lake City	TRUE
Los Angeles	Hartford-New Haven	TRUE
Philadelphia	Austin	TRUE
Salt Lake City		

As of now, our startup locations are narrowed down to Austin, Chicago, San Diego, Denver, Los Angeles, Philadelphia, and Salt Lake City. As mentioned above these locations- while not the top performing based on average weekly sales, still hold strong. In fact, it may be ideal to select locations where competitors are not performing best, and the overall criteria are met. This would mean that potential sales still remain for the startup.

Further analysis at our income based locations shows us that the average household income of Houston is just shy of our \$65K at \$64,432. Given that Houston is the highest performing location in the competitor data, we will go ahead and include it into our startup locations. While selecting locations where competitors aren't as strong is smart it's also important to make sure there exists locations in competitor dominant areas.

As stated by the blog by Cornell University, there exists this notion that business attempt to setup locations near the strongest "hubs" of the population. Doing so means that they are poised to attract the largest number of customers. In addition, to this, it has been shown that when smaller companies open locations to known brands it creates an association in the shopper's perspective of brand strength. That is, our startup will become associated with being a competitive and viable alternative to our larger more established competitors. This provides a potential shortcut in establishing some form of brand strength. The phenomenon described above is known as "Hotelling's Model of Spatial Competition".

All of this is to say that we are now left with 2 additional locations to add to our list. Just as described above, we will make an exception to the strict income barrier of \$65K, and select the remaining top 2 performing competitor locations found in the data. Those locations being Orlando-Daytona Beach-Melbourne and Tampa-St Petersburg.

In total our finalized startup locations list is as follows: **Austin, Chicago, San Diego, Denver, Los Angeles, Philadelphia, Salt Lake City, Orlando-Daytona Beach-Melbourne, and Tampa-St Petersburg.** We've selected these locations based on income, population, and competitor performance. We've also taken into consideration foundational business practices in our choice selection. Selecting these 10 locations will provide a strong backbone in the startup's launch and future.

Resources:

1. The Ifod - [link](#)
2. Cornell University - [link](#)
3. Investopedia - [link](#)
4. Kaggle Dataset - [link](#)